Jeu de coalition pour la décharge des taches dans les infrastructures Fog

(Rapport de Classe)

Asma El Kounti

Abstract

Les tâches de calcul comme le contenu multimédia produisent du contenu qui peut être décrit, stocké, mis en cache. Dans cet article, nous étudions le problème de l'informatique de périphérie et du déchargement dans les réseaux d'informatique en fog. Les dispositifs des utilisateurs ont toujours été limités dans leurs capacités de calcul. Le problème de la latence se pose régulièrement lorsqu'il y a une grande quantité de données à traiter. En considérant un nombre donné d'infrastructures fog servant plusieurs utilisateurs, cet article étudie le problème de la distribution des tâches de calcul entre les infrastructures fog. Nous proposons une solution basée sur un jeu de coalition [1] pour motiver les fogs à travailler ensemble et à coopérer, ce qui permet de traiter d'énormes quantités de données dans un délai réduit. Les résultats de la simulation montrent que le modèle proposé offre une meilleure latence et un meilleur équilibrage de la répartition des ressources des fogs.

Index Terms

Fog Computing, Jeu de Coalition, Collaboration Fog-to-Fog

I. Introduction

Ce monde est plein d'innovations numériques où une variété d'informations et d'objets sont connectés les uns aux autres sur différents réseaux. Les objets sont coordonnés ensemble et coordonnés avec leurs informations respectives. Lorsqu'une variété d'informations est connectée à l'internet, on l'appelle l'internet des objets (IoT) [2]–[4]. De nouvelles applications et de nouveaux dispositifs sont appliqués en ligne pour mesurer, surveiller, traiter, analyser et réagir à une tempête de données apparemment sans fin. Le IoT est un régime dans lequel nous sommes entourés de milliards d'appareils intelligents connectés qui influencent une partie de la vie humaine [5]–[7].

Les données extrêmement variées et incomplètes générées par l'IoT doivent être traitées et exploitées dans un délai très réduit. Aujourd'hui, le cloud est devenu un élément indispensable de ce processus [8]–[10]. Cependant, le cloud a été déployé de manière centralisée à l'échelle mondiale. Le cloud doit traiter une grande quantité de données. En outre, plus la distance physique entre l'utilisateur et le cloud augmente, plus la latence de transmission augmente. Cela augmente le temps de réponse et frustre l'utilisateur. Le cloud computing est largement utilisé. Il présente certaines limites en raison des grandes quantités de données. La limitation fondamentale est la connectivité entre le cloud et les dispositifs finaux, selon [11]–[13]. Cette connectivité est établie sur Internet, ce qui ne convient pas à un grand nombre d'applications basées sur le cloud computing, comme celles qui sont sensibles à la latence. En outre, les applications basées sur le cloud sont souvent distribuées et comportent de multiples composants [14]–

[17]. Il n'est donc pas rare que des composants d'applications soient déployés séparément sur plusieurs clouds. Ce phénomène peut aggraver la latence en raison de la surcharge introduite par les communications inter-clouds.

Le Fog Computing accélère le traitement des données, tout en fournissant simultanément la connaissance de l'emplacement et la faible latence en éliminant tout le parcours jusqu'aux clouds à distance [18], [19]. La connaissance de l'emplacement de la couche de fog peut être utilisée pour répondre à de multiples exigences fonctionnelles et non fonctionnelles des applications, notamment la mobilité et la sécurité. Le fog donne au cloud la capacité de traiter les énormes données générées par les appareils et l'IoT. Il s'agit d'une autre architecture distribuée qui représente un élément de continuité entre le cloud et le reste des composant réseaux. Le concept de fog est le dernier descendant en date de la séparation physique des unités fonctionnelles [20]. Il s'agit d'une couche informatique plus proche de la couche de perception, où résident les capteurs et les déclencheurs, et qui fournit des services de calcul, de mise en réseau et de stockage. Cette proximité offre une série d'avantages qui caractérisent cette couche [21].

Dans [2], les auteurs ont formulé un jeu de déchargement de calcul pour structurer la concurrence entre les utilisateurs de l'IoT et allouer efficacement la puissance de traitement limitée des nœuds de fog. Chaque utilisateur a pour objectif de maximiser sa propre qualité d'expérience. Les auteurs proposent une allocation de ressources quasi-optimale dans un paradigme d'informatique en cloud et de fog hiérarchique et prouvent que dans un système avec N utilisateurs IoT, il peut atteindre un équilibre de Nash. La qualité d'expérience des utilisateurs étant évaluée ainsi que l'efficacité de l'équilibre, les auteurs ont formulé un problème de maximisation de la qualité d'expérience basé sur un jeu de déchargement de calcul pour structurer la concurrence entre les utilisateurs IoT. Le résultat est qu'un plus grand nombre d'utilisateurs profitent des services informatiques et du délai de calcul, ce qui permet d'offrir des services informatiques en fog à faible latence pour les applications IoT sensibles aux délais. Dans [22], les auteurs explorent la question de la performance du délai, de la consommation d'énergie et du coût du paiement dans un système mobile d'informatique en fog. Ils ont optimisé la probabilité de déchargement et la puissance de transmission pour les dispositifs mobiles afin de minimiser les performances de retard, la consommation d'énergie et le coût. Les résultats analytiques ont été mesurés sur la consommation d'énergie, la performance de retard et le coût de paiement. Les auteurs de [23] ont proposé un ordonnancement prédictif dans un réseau de mise en cache sans fil pour le problème de l'association utilisateur-AP et de l'allocation dynamique des ressources. En utilisant l'information sur la structure du problème, ils proposent un algorithme en ligne appelé prédictive. Cet algorithme en ligne garantit la performance et ne nécessite aucune information statistique du système dynamique. Dans [16], [24], [25], les auteurs ont proposé un schéma pour combiner l'association des utilisateurs, le taux de livraison des données et le pré-codage du signal dans la liaison descendante d'un F-RAN avec cache et une capacité frontale limitée, ils ont formulé un problème visant à maximiser la différence pondérée du débit du réseau et de la consommation d'énergie totale.

Dans cet article, nous étudions le problème des tâches informatiques distribuées entre les fog. Nous introduisons ensuite un cadre général de bon sens pour comprendre, évaluer et modéliser le retard de service dans les scénarios d'application de fog à fog essayant de travailler ensemble pour traiter d'énormes quantités de données dans un court délai tels que dans [25]–[27]. Nous proposons ensuite une affectation des tâches entre les fog. Nous proposons dans

notre scénario, un modèle théorique d'affectation des tâches entre les fog. Nous proposons dans notre scénario, une solution théorique basée sur un jeu, un jeu de coalition. L'objectif est d'optimiser les ressources des fog et d'offrir une meilleure qualité de service (QoS) et une meilleure expérience pour les utilisateurs (QoE). De plus, notre schéma n'est pas limité à une architecture spécifique (comme un réseau cellulaire), au nombre, au type ou à la topologie des nœuds IoT, des nœuds de fog et des serveurs cloud. Les résultats des simulations montrent que la performance du modèle proposé par rapport à un modèle de coalition normal et une politique d'affectation aléatoire offre et garantit la meilleure QoS/E et l'optimisation des ressources.

Ce document est organisé comme suit. La deuxième section décrit les différentes entités du système réseaux, ainsi que les différents paramètres. Dans la troisième section nous formulons le problème de décharge des taches de traitement. La solution proposée est décrite dans la quatrième section. Les simulations et évaluation de notre approche sont présenté dans la section cinq. Enfin, nous concluons le papier en sixième section.

II. System Model

Afin de mieux répondre aux demandes des utilisateurs, de réduire la latence et d'optimiser l'équilibrage de la charge entre les nœuds de brouillard, nous considérons un réseau constitué d'un ensemble de brouillards $\mathcal{F}$ de F fogs répartis sur une zone, $\mathcal{F} = f_1, f_2, ..., f_F$. Chaque fog possède ce que nous appelons le vecteur d'état, c'est un vecteur contenant des informations sur les ressources du fog (CPU, stockage, mémoire, ...), et nous supposons que les brouillards partagent leur vecteur d'état entre eux. Un ensemble d'utilisateurs, qui peuvent être des capteurs ou des utilisateurs mobiles. Ces utilisateurs sont répartis sur la sous-zone couverte par les brouillards (voir fig. 1). Nous considérons que les utilisateurs sont associés aléatoirement aux fogs, et nous considérons la variable binaire $x_{(u,f)}$ qui exprime si un utilisateur u est associé au brouillard f_i. On dit que $x_{(u,f)} = 1$ si un utilisateur u est associé à un fog f, et égale à 0 au cas contraire. Le nombre d'utilisateurs associés d'un fog particulier correspond à la somme des éléments suivants $x_{(u,f)}$, définie comme suit :

$$N^f = \sum_{u \in \mathcal{U}} x_{u,f} \tag{1}$$

Nous considérons que la taille des tâches suit une distribution exponentielle de moyenne ω et que l'arrivée des demandes suit une distribution de Poison de moyenne λ. Lorsqu'un utilisateur demande une tâche de traitement au fog qui lui est associé, le temps de réponse dépend de la taille des tâches, de la qualité du canal de communication et du temps d'attente. Le délai total qu'une demande subira est donné par (d'après [10], [15], [16]):

$$L_u(t) = \frac{\omega_u}{F_i} + \delta_Q + \delta_P \tag{2}$$

Où ω_u est la taille de la tâche de l'utilisateur u, c_f est la capacité de traitement du fog f, δ_Q et δ_P sont respectivement le temps d'attente avant d'être servi et le temps qu'une requête prend pour arriver au fog puis retourner à l'utilisateur (dépend de la qualité du canal), nous considérons que les canaux sont fiables et nous négligeons l'impact de cette variable (i.e : $\tau_P = 0$). Le temps de file d'attente est donné par :

Fig. 1: Architecture du système

$$\delta_Q = \sum_{u \in \mathcal{U}} \frac{\omega_u}{F_i} \tag{3}$$

Où ω_u est la taille de la tâche de l'utilisateur i et Cf est la capacité maximale du fog f. Lorsqu'un fog donné reçoit une tâche de traitement lourde, il obtient l'état des autres fogs et essaie de répartir les tâches entre eux. En distribuant le traitement, nous pouvons améliorer la latence. Pour simplifier le modèle, nous considérons que lorsqu'une tâche est distribuée, nous divisons le temps par le nombre de nœuds participants (ex : si nous avons 5 nœuds, le délai de traitement sera taille de la tâche / capacités de traitement / 5). L'objectif principal de ce papier est de réduire la latence des requêtes, en profitant de la communication horizontale entre les fogs. A cette fin, nous proposons une approche basée sur un jeu de coalition pour motiver les fogs à collaborer entre eux afin d'offrir une meilleure latence aux utilisateurs.

$$x_{i,j} = \begin{cases} 1 \text{ si } j \text{ est associé à } i \\ 0 \text{ Sinon} \end{cases} \tag{4}$$

III. Jeu de Coalition pour décharge des taches

Nous présentons dans cette section la formulation de solution à base de la thèorie de jeu. Afin d'obtenir une meilleure latence, l'utilisation de la proximité des fogs entre eux pourrait réduire la latence. Dans notre modèle, nous supposons que les fogs peuvent communiquer horizontalement. L'idée est que lorsqu'un utilisateur donné demande une tâche de traitement, le fog associé vérifie l'état de ses ressources. S'il peut traiter la tâche, il la traite. Sinon, il peut vérifier le vecteur d'état de ses voisins et déléguer une partie des tâches à un ou plusieurs fogs. Figure 2. Pour chaque tâche reçue, le fog estime la latence en utilisant son vecteur d'état et la taille de la tâche. Il étiquette également les tâches qui seront distribuées.

Le jeu est formulé comme suit :

- L'utilisateur rejoint les coalitions
- Le nombre de coalitions est le nombre de fogs.
- Les utilisateurs rejoignent la coalition en fonction de celle qui offre la plus faible latence.

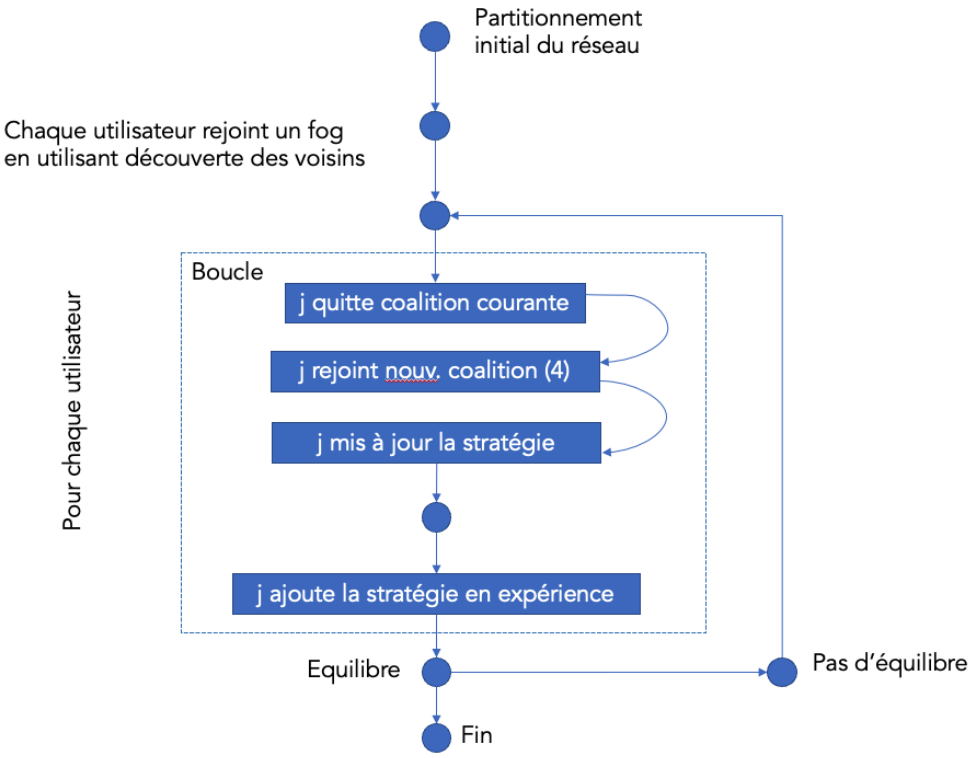


Fig. 2: L'approche proposé

Le nombre d'utilisateur pour chaque coalision est donné par,

$$l_i = \sum_{j=0}^{j=M} x_{i,j} \tag{5}$$

La fonction d'utilité est consideré comme étant la fonction $\mathcal{L}_i(t)$,

$$\mathcal{L}_i(t) = \frac{d_j}{F_i} + \sum_{\substack{j'=0 \\ j \neq j'}}^{M-1} \frac{d'_j}{F_i} \tag{6}$$

Le problème est formulé comme suit,

$$\text{minimize} \quad \mathcal{L}_i \tag{7a}$$

$$\text{subject to} \quad \sum_{j=0}^{j=M} x_{i,j} \geq 1, \tag{7b}$$

$$x_{i,j} \in 0, 1 \tag{7c}$$

L'équilibre de Nash est atteint lorsqu'aucun membre d'une coalition ne veut changer de stratégie. membre d'une coalition ne veut changer sa stratégie, ce qui va évidemment la performance, c'est-à-dire la latence pour les utilisateurs et les capacités d'utilisation des brouillards. l'utilisation des capacités des fogs. L'équilibre de Nash est exprimé comme suit,

$$U_i(\sigma_i^*, \sigma_{-i}^*)) \geq U_i(\sigma_i, \sigma_{-i}^*)) \tag{8}$$

où σ_i est la stratégie utilisée par le joueur i, σ_i^* est la stratégie utilisée par tous les joueurs sauf le joueur i et U_i est la fonction de gain qui représente les performances du joueur i. L'inégalité signifie qu'aucun joueur ne peut faire mieux en s'écartant unilatéralement de l'équilibre. La mise à jour de la distribution des utilisateurs est effectuée à chaque instant t, pour un paramètre donné, [1].

$$\omega_i(t) = \frac{\omega_u(t-1)(1+\epsilon)^{P_u}}{\sum_i \omega_i(t-1)(1+\epsilon)^{P_i}} \tag{9}$$

où P_i est le profit que la ième stratégie recevrait au temps t. Pendant la phase de clustering, nous considérons que l'opération commence à la réception de la deuxième requête, et nous utiliserons la première comme pivot. De plus, le nombre de clusters est connu, puisque nous associons les utilisateurs aux fogs [c-à-d : chaque fog représente un cluster].

IV. Simulations et Experimentations

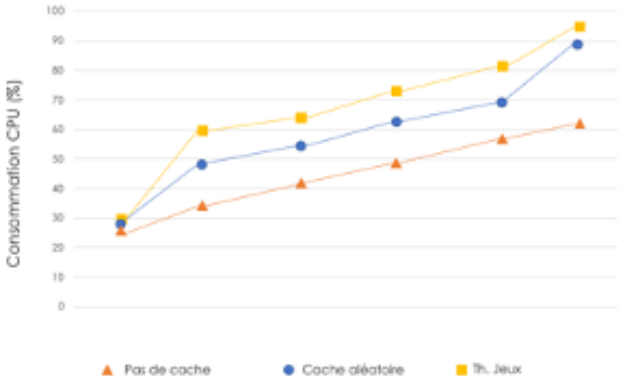
Nous avons évalué le modèle proposé en utilisant 4 fogs pour tester l'utilisation des ressources et 5 fogs pour tester la latence moyenne. Nous considérons trois scénarios, notre modèle proposé, une politique de mise en cache aléatoire et le cas où aucun schéma n'a été considéré. En appliquant notre approche, la consommation de ressources de chaque fog est améliorée. En fait, la distribution des tâches dans les autres fogs donne une meilleure latence en raison de la distribution du traitement. La figure 3 montre que l'utilisation des ressources (traitement) de certains fogs est passée de 95 % à 69 %. Ce qui aura un impact sur l'équilibrage de la charge. La figure 4 montre la latence moyenne du même ensemble de tâches dans les mêmes conditions et dans différents schémas. Dans la plupart des cas de simulations, il est montré que la latence moyenne de notre schéma est la meilleure La latence moyenne est réduite (de 200 ms à moins de 150 ms dans certains cas).

V. Conclusion

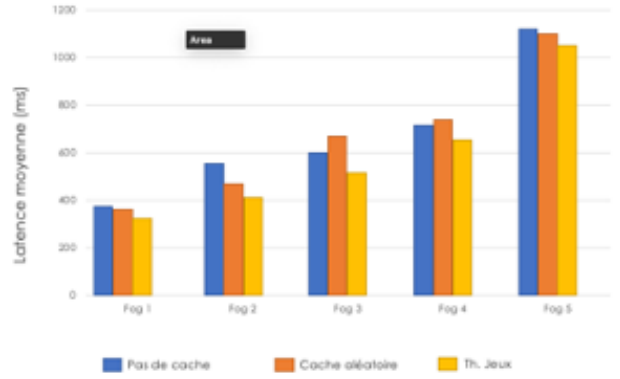
Dans ce travail, nous avons abordé le problème de l'informatique de périphérie et du déchargement dans les réseaux d'informatique en nuage. Les dispositifs de périphérie ont toujours été limités dans leurs capacités de calcul. Le problème de latence se pose fréquemment lorsqu'il y a une énorme quantité de données à traiter. En considérant un nombre donné de fogs desservant plusieurs utilisateurs, cet article étudie le problème de la distribution des tâches de calcul entre les infrastructures fog. Nous avons proposé une solution basée sur les jeux de coalition pour motiver les fogs à coopérer, ce qui a permis de traiter d'énormes quantités de données dans un court délai. Les résultats de la simulation primaire montrent que le modèle proposé offre une latence meilleure et consomme moins de CPU.

References

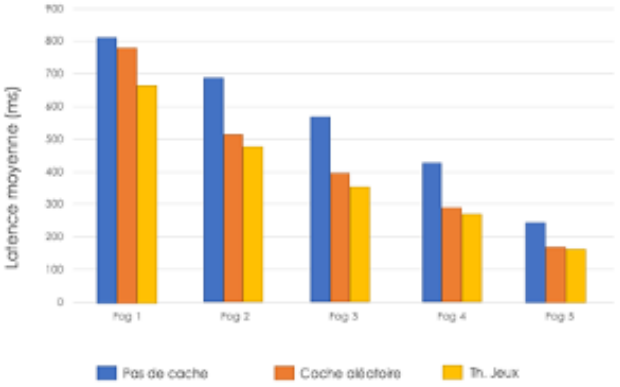
[1] C. Daskalakis *et al.*, "On learning algorithms for nash equilibria," in *International Symposium on Algorithmic Game Theory*. Springer, 2010, pp. 114–125.



(a) La latence en fonction de la configuration des fogs



(b) La latence en fonction de la taille des tâches à décharger



(c) Consommation CPU en fonction de la taille de requêtes

Fig. 3: Simulation results.

[2] L. Liu *et al.*, "Multi-objective optimization for computation offloading in fog computing," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 283–294, 2017.

[3] M.-A. Koulali *et al.*, "Qdgrp : A hybrid qos distributed genetic routing protocol for wireless sensor networks," in *2012 International Conference on Multimedia Computing and Systems*, 2012, pp. 47–52.

[4] A. Harbouche *et al.*, "A flexible wireless body sensor network system for health monitoring," in *2013 Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*, 2013, pp. 44–49.

[5] M.-A. Koulali *et al.*, "Industrial internet of things-based prognostic health management: a mean-field stochastic game approach," *IEEE Access*, vol. 6, pp. 54 388–54 395, 2018.

[6] A. Walid *et al.*, "A decentralized network selection algorithm for group vertical handover in heterogeneous networks," in *2014 IEEE wireless communications and networking conference (WCNC)*. IEEE, 2014, pp. 2817–2821.

[7] A. Walid *et al.*, "Group vertical handoff management in heterogeneous networks," *Wireless Communications and mobile computing*, vol. 16, no. 10, pp. 1256–1270, 2016.

[8] A. Mabrouk *et al.*, "Meeting always-best-connected paradigm in heterogeneous vehicular networks: A graph theory and a signaling game analysis," *Vehicular Communications*, vol. 5, pp. 1–8, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214209616300328

[9] A. Walid *et al.*, "A decentralized network selection algorithm for group vertical handover in heterogeneous networks," in *2014 IEEE Wireless Communications and Networking Conference (WCNC)*, 2014, pp. 2817–2821.

[10] O. Ait Oualhaj *et al.*, "A coalitional-game-based incentive mechanism for content caching in heterogeneous delay tolerant networks," in *2015 International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2015, pp. 987–992.

[11] M. Amine *et al.*, "Matching game for green uplink in hyper dense lte hetnets," in *2016 International Conference on Wireless Networks and Mobile Communications (WINCOM)*, 2016, pp. 133–139.

[12] M. Amine *et al.*, "A many-to-many matching game in ultra-dense lte hetnets," in *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2017, pp. 1245–1250.

[13] A. Abouaomar *et al.*, "Caching, device-to-device and fog computing in 5th cellular networks generation : Survey," in *2017 International Conference on Wireless Networks and Mobile Communications (WINCOM)*, 2017, pp. 1–6.

[14] M. Amine *et al.*, "New user association scheme based on multi-objective optimization for 5g ultra-dense multi-rat hetnets," in *2018 IEEE International Conference on Communications (ICC)*, 2018, pp. 1–6.

[15] A. Abouaomar *et al.*, "Users-fogs association within a cache context in 5g networks:coalition game model," in *2018 IEEE Symposium on Computers and Communications (ISCC)*, 2018, pp. 00 014–00 019.

[16] Z. Ennya *et al.*, "Computing tasks distribution in fog computing: Coalition game model," in *2018 6th International Conference on Wireless Networks and Mobile Communications (WINCOM)*, 2018, pp. 1–4.

[17] A. Abouaomar *et al.*, "Matching-game for user-fog assignment," in *2018 IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 1–6.

[18] A. Abouaomar *et al.*, "A resources representation for resource allocation in fog computing networks," in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.

[19] M. Amine *et al.*, "New network slicing scheme for ue association solution in 5g ultra dense hetnets," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[20] A. Filali *et al.*, "Multi-access edge computing: A survey," *IEEE Access*, vol. 8, pp. 197 017–197 046, 2020.

[21] S. Brihi *et al.*, "A new profile-qos aware user association scheme for 5g ultra-dense mmw hetnets," in *2018 IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 1–6.

[22] T. T. Vu *et al.*, "Joint optimization of user association, data delivery rate and precoding for cache-enabled f-rans," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*. IEEE, 2017, pp. 1–6.

[23] S. Zhao *et al.*, "Online user-ap association with predictive scheduling in wireless caching networks," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*. IEEE, 2017, pp. 1–7.

[24] M. Amine *et al.*, "Two-sided matching framework for optimal user association in 5g multi-rat udns," *International Journal of Communication Systems*, vol. 33, no. 8, p. e4346, e4346 dac.4346. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/dac.4346

[25] A. Abouaomar *et al.*, "Mean-field game and reinforcement learning mec resource provisioning for sfc," in *2021 IEEE Global Communications Conference (GLOBECOM)*, 2021, pp. 1–6.

[26] M. Elmachkour *et al.*, "Data traffic-based analysis of delay and energy consumption in cognitive radio networks with and without resource reservation," *International Journal of Communication Systems*, vol. 28, no. 7, pp. 1316–1328, 2015.

[27] M. Elmachkour *et al.*, "The greening of spectrum sensing: A minority game-based mechanism design," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 150–156, 2014.

[28] S. Brihi *et al.*, "Uav for iot communications: Beam selection using matching game and network slicing," in *2020 8th International Conference on Wireless Networks and Mobile Communications (WINCOM)*, 2020, pp. 1–7.

[29] O. Ait Oualhaj *et al.*, "A ferry-assisted solution for forwarding function in wireless sensor networks," *Pervasive and Mobile Computing*, vol. 22, pp. 126–135, 2015, special Issue on Recent Developments in Cognitive Radio Sensor Networks. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1574119215000905

[30] A. Abouaomar *et al.*, "A deep reinforcement learning approach for service migration in mec-enabled vehicular networks," in *2021 IEEE 46th Conference on Local Computer Networks (LCN)*, 2021, pp. 273–280.

[31] A. Abouaomar *et al.*, "Resource provisioning in edge computing for latency-sensitive applications," *IEEE Internet of Things Journal*, vol. 8, no. 14, pp. 11 088–11 099, 2021.

[32] H. Shah-Mansouri *et al.*, "Hierarchical fog-cloud computing for iot systems: A computation offloading game," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 3246–3257, 2018.